(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 02/103673 A1

(51) International Patent Classification[7]: G10L 15/00, G06F 3/16, 17/20, 17/27, 17/28

(21) International Application Number: PCT/AU02/00803

(22) International Filing Date: 19 June 2002 (19.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PR 5793 19 June 2001 (19.06.2001) AU

(71) Applicant *(for all designated States except US)*: **KAZ GROUP LIMITED** [AU/AU]; Level 7, 66 Wentworth Avenue, Sydney, NSW 2010 (AU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TALHAMI, Habib** [AU/AU]; Level 7, 66 Wentworth Avenue, Sydney, NSW 2010 (AU). **WALDRON, Nik** [AU/AU]; Level 7, 66 Wentworth Avenue, Sydney, NSW 2010 (AU).

(74) Agent: **WALLINGTON-DUMMER**; P.O. Box 297, Rydalmere, NSW 1701 (AU).

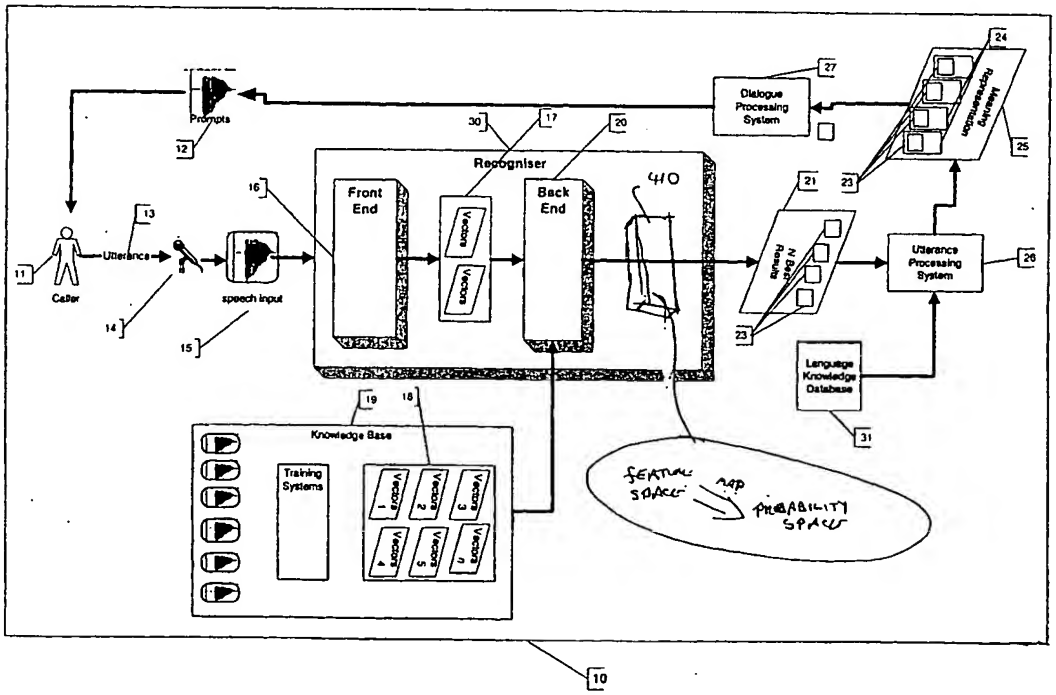
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *with international search report*
— *with amended claims*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: NEURAL NETWORK POST-PROCESSOR




(57) Abstract: In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of gauging correctness of a pattern recognition task by mapping an input feature space onto a probability space.

WO 02/103673 A1

# NEURAL NETWORK POST-PROCESSOR

The present invention relates to a neural network post-processor and more particularly to such a processor incorporated within a recogniser portion of an automated speech recognition system.

## BACKGROUND

Automated speech recognition is a complex task in itself. Automated speech understanding sufficient to provide automated dialogue with a user adds a further layer of complexity.

In this specification the term "automated speech recognition system" will refer to automated or substantially automated systems which perform automated speech recognition and also attempt automated speech understanding, at least to predetermined levels sufficient to provide a capability for at least limited automated dialogue with a user.

A generalized diagram of a commercial grade automated speech recognition system as can be used for example in call centres and the like is illustrated in Fig. 1.

With advances in digital computers and a significant lowering in cost per unit of computing capacity there have been a number of attempts in the commercial marketplace to install such automated speech recognition systems implemented substantially by means of digital computers.

However, to date, there remain problems in achieving 100% recognition and/or 100% understanding in real time.

In one particular form critical to the success or otherwise of any given recognition schema there are difficulties in classifying patterns as correctly recognized or incorrectly recognized/not modeled.

It is an object of the present invention to address or ameliorate one or more of the abovementioned disadvantages.

## BRIEF DESCRIPTION OF INVENTION

Accordingly, in one broad form of the invention there is provided in a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of gauging correctness of a pattern recognition task by mapping an input feature space onto a probability space.

Preferably said mapping is a non-linear mapping.

Preferably said method is applied to a confidence scoring task.

Preferably said method utilizes a multi-layer perceptron to apply multiple knowledge sources to said confidence scoring task.

In a further broad form of the invention there is provided in a speech recognition system of the type adapted

to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of obtaining a confidence score by a non-linear mapping onto the real number line.

Preferably said method utilizes an MLP to generate an aposteriori probability for confidence.

Preferably said MLP is trained with a mean squared error.

Preferably said MLP is trained utilizing a cross-entropy cost function.

Preferably said MLP is additionally trained with some sigmoidal non-linearity.

In yet a further broad form of the invention there is provided in a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of confidence scoring utilizing a data driven system.

**BRIEF DESCRIPTION OF DRAWINGS**

Embodiments of the present invention will now be described with reference to the accompanying drawings wherein:

Fig. 1 is a generalized block diagram of a prior art automated speech recognition system;

Fig. 2 is a generalized block diagram of an automated speech recognition system suited for use in conjunction with an embodiment of the present invention;

Fig. 3 is a more detailed block diagram of the utterance processing and dialogue processing portions of the system of Fig. 2;

Fig. 4 is a block diagram of the system of Fig. 2 incorporating a neural network post-processor in accordance with a first embodiment of the present invention.

## DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

With reference to Fig. 2 there is illustrated a generalized block diagram of an automated speech recognition system 10 adapted to receive human speech derived from user 11, and to process that speech with a view to recognizing and understanding the speech to a sufficient level of accuracy that a response 12 can be returned to user 11 by system 10. In the context of systems to which embodiments of the present invention are applicable the response 12 can take the form of an auditory communication, a written or visual communication or any other form of communication intelligible to user 11 or a combination thereof.

In all cases input from user 11 is in the form of a plurality of utterances 13 which are received by transducer 14 (for example a microphone) and converted into an electronic representation 15 of the utterances 13. In one exemplary form the electronic representation 15 comprises a digital representation of the utterances 13 in .WAV format. Each electronic representation 15 represents an entire utterance 13. The electronic representations 15 are processed through front end processor 16 to produce a stream of vectors 17, one vector for example for each 10ms segment of utterance 13. The vectors 17 are matched against knowledge base vectors 18 derived from knowledge base 19 by back end processor 20 so as to produce ranked results 1-N in the form of N best results 21. The results can comprise for example subwords, words or phrases but will depend on the application. N can vary from 1 to very high values, again depending on the application.

An utterance processing system 26 receives the N best results 21 and begins the task of assembling the results into a meaning representation 25 for example based on the data contained in language knowledge database 31.

The utterance processing system 26 orders the resulting tokens or words 23 contained in N-best results 21 into a meaning representation 25 of token or word candidates which are passed to the dialogue processing system 27 where sufficient understanding is attained so as

to permit functional utilization of speech input 15 from user 11 for the task to be performed by the automated speech recognition system 10. In this case the functionality includes attaining of sufficient understanding to permit at least a limited dialogue to be entered into with user/caller 11 by means of response 12 in the form of prompts so as to elicit further speech input from the user 11. In the alternative or in addition, the functionality for example can include a sufficient understanding to permit interaction with extended databases for data identification.

Fig. 3 illustrates further detail of the system of Fig. 2 including listing of further functional components which make up the utterance processing system 26 and the dialogue processing system 27 and their interaction. Like components are numbered as for the arrangement of Fig. 2.

The utterance processing system 26 and the dialogue processing system 27 together form a natural language processing system. The utterance processing system 26 is event-driven and processes each of the utterances 13 of caller/user 11 individually. The dialogue processing system 27 puts any given utterance 13 of caller/user 11 into the context of the current conversation (usually in the context of a telephone conversation). Broadly, in a telephone answering context, it will try to resolve the

query from the caller and decide on an appropriate answer to be provided by way of response 12.

The utterance processing system 26 takes as input the output of the acoustic or speech recogniser 30 and produces a meaning representation 25 for passing to dialogue processing system 27.

In a typical, but not limiting form, the meaning representation 25 can take the form of value pairs. For example, the utterance "I want to go from Melbourne to Sydney on Wednesday" may be presented to the dialogue processing system 27 in the form of three value pairs, comprising:

1. Start; Melbourne
2. Destination; Sydney
3. Date; Wednesday

where, in this instance, the components Melbourne, Sydney, Wednesday of the value pairs 24 comprise tokens or words 23.

With particular reference to Fig. 3 the recogniser 30 provides as output N best results 21 usually in the form of tokens or words 23 to the utterance processing system 26 where it is first disambiguated by language model 32. In one form the language model 32 is based on trigrams with cut off.

Analyser 33 specifies how words derived from language model 32 can be grouped together to form meaningful phrases

which are used to interpret utterance 13. In one form the analyzer is based on a series of simple finite state automata which produce robust parses of phrasal chunks - for example noun phrases for entities and concepts and WH-phrases for questions, dates. Analyser 33 is driven by grammars such as meta-grammar 34. The grammars themselves must be tailored for each application and can be thought of as data created for a specific customer.

The resolver 35 then uses semantic information associated with the words of the phrases recognized as relevant by the analyzer 33 to refine the meaning representation 25 into its final form for passing through the dialogue flow controller 36 within dialogue processing system 27.

The dialogue processing system 27, in this instance with reference to Fig. 3, receives meaning representation 25 from resolver 35 and processes the dialogue according to the appropriate dialogue models. Again, dialogue models will be specific to different applications but some may be reusable. For example a protocol model may handle greetings, closures, interruptions, errors and the like across a number of different applications.

The dialogue flow controller 36 uses the dialogue history to keep track of the interactions.

The logic engine 37, in this instance, creates SQL queries based on the meaning representation 25. Again it

will be dependent on the specific application and its domain knowledge base.

The generator 38 produces responses 12 (for example speech out). In the simplest form the generator 38 can utilize generic text to speech (TTS) systems to produce a voiced response.

Language knowledge database 31 comprises, in the instance of Fig. 3, a lexicon 39 operating in conjunction with database 40. The lexicon 39 and database 40 operating in conjunction with knowledge base mapping tools 41 and, as appropriate, language model 32 and grammars 34 constitutes a language knowledge database 31 or knowledge base which deals with domain specific data. The structure and grouping of data is modeled in the knowledge base 31.

Database 40 comprises raw data provided by a customer. In one instance this data may comprise names, addresses, places, dates and is usually organised in a way that logically relates to the way it will be used. The database 40 may remain unchanged or it may be updated throughout the lifetime of an application. Functional implementation can be by way of database servers such as MySQL, Oracle, Postgres.

As will be observed particularly with reference to Fig. 3, interaction between a number of components in the system can be quite complex with lexicon 39, in particular,

being used by and interacting with multiple components of System 10.

With reference to Fig. 4 there is shown in block diagram form a neural network post-processor 410 in accordance with a first preferred embodiment of the present invention.

In this instance the processor 410 is applied to the output of recogniser 30.

Broadly neural network post-processor 410 utilises a multi-layer perceptron to apply multiple knowledge sources to the problem. This performs a non-linear mapping onto the real number line to give us a confidence score.

This differs from previous solutions in several ways:

1. The application of multiple knowledge sources
2. The use of an MLP to generate an aposteriori probability for confidence.

This solution can be implemented simply using an MLP trained with either a mean squared error or a cross-entropy cost function, and some sigmoidal non-linearity. Our experimental system was trained using conjugate gradient descent (back-propagation).

The system 10 incorporating processor 410 is data driven, rather than based on some heuristic technique, as such (for a representative corpus). It 'learns' an optimal mapping from input data to a correct/incorrect mapping.

In effect, in order to gauge end best results 21 derived from recogniser 30 the neural network post-processor 410 non-linearly maps an input feature space of a pattern recognition task onto the probability space for gauging correctness as applied to confidence scoring.

The above describes only some embodiments of the present invention and modifications, obvious to those skilled in the art, can be made thereto without departing from the scope and spirit of the present invention.

## CLAIMS

1. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of gauging correctness of a pattern recognition task by mapping an input feature space onto a probability space.

2. The method of Claim 1 wherein said mapping is a non-linear mapping.

3. The method of Claim 1 or Claim 2 applied to a confidence scoring task.

4. The method of any previous claim utilizing a multi-layer perceptron to apply multiple knowledge sources to said confidence scoring task.

5. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of obtaining a confidence score by a non-linear mapping onto the real number line.

6. The method of any previous claim utilizing an MLP to generate an aposteriori probability for confidence.

7. The method of Claim 6 wherein said MLP is trained with a mean squared error.

8. The method of Claim 6 wherein said MLP is trained utilizing a cross-entropy cost function.

9. The method of Claim 7 and Claim 8 wherein said MLP is additionally trained with some sigmoidal non-linearity.

10. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of confidence scoring utilizing a data driven system.

11. A speech recognition system operating according to the method of any previous claim.

## AMENDED CLAIMS

[received by the International Bureau on 11 November 2002 (11.11.02);
original claims 1-11 replaced by new claims 1-17]

1. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of gauging correctness of a pattern recognition task by mapping an input feature space onto a probability space.

2. The method of Claim 1 wherein said mapping is a non-linear mapping.

3. The method of Claim 1 or Claim 2 applied to a confidence scoring task.

4. The method of any previous claim utilizing a multi-layer perceptron (MLP) to apply multiple knowledge sources to said confidence scoring task.

5. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of obtaining a confidence score by a non-linear mapping onto the real number line.

6. The method of any previous claim utilizing an MLP to generate an aposteriori probability for confidence.

7. The method of Claim 6 wherein said MLP is trained with a mean squared error.

8. The method of Claim 6 wherein said MLP is trained utilizing a cross-entropy cost function.

9. The method of Claim 7 and Claim 8 wherein said MLP is additionally trained with some sigmoidal non-linearity.

10. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of confidence scoring utilizing a data driven system.

11. A speech recognition system operating according to the method of any previous claim.

12. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of gauging correctness of a pattern recognition task by mapping an input feature

space onto a probability space; and wherein said mapping is a non-linear mapping; and wherein said method is applied to a confidence scoring task; said method utilizing a multi-layer perceptron to apply multiple knowledge sources to said confidence scoring task; said method utilizing said multi-layer perceptron to generate an aposteriori probability for confidence and wherein said multi-layer perceptron is trained with a mean squared error; said multi-layer perceptron also trained utilizing a cross-entropy cost function.

13. The method of Claim 12 utilising a data driven system.

14. The method of Claim 12 or 13 wherein said multi-layer perceptron is additionally trained with some sigmoidal non-linearity.

15. A speech recognition system operating according to the method of any one of Claims 12 to 14.

16. In a speech recognition system of the type adapted to process utterances from a caller or user by way of a recogniser, an utterance processing system and a dialogue processing system so as to produce responses to said utterances, a method of gauging correctness of a pattern recognition task by mapping an input feature space onto a probability space; and wherein said

mapping is a non-linear mapping; and wherein said method is applied to a confidence scoring task; said method utilizing a multi-layer perceptron to apply multiple knowledge sources to said confidence scoring task; said method utilizing said multi-layer perceptron to generate an aposteriori probability for confidence and wherein said multi-layer perceptron is trained with a mean squared error; said multi-layer perceptron also trained utilizing a cross-entropy cost function; and wherein said multi-layer perceptron is additionally trained with some sigmoidal non-linearity.

17. A speech recognition system operating according to the method of Claim 16.

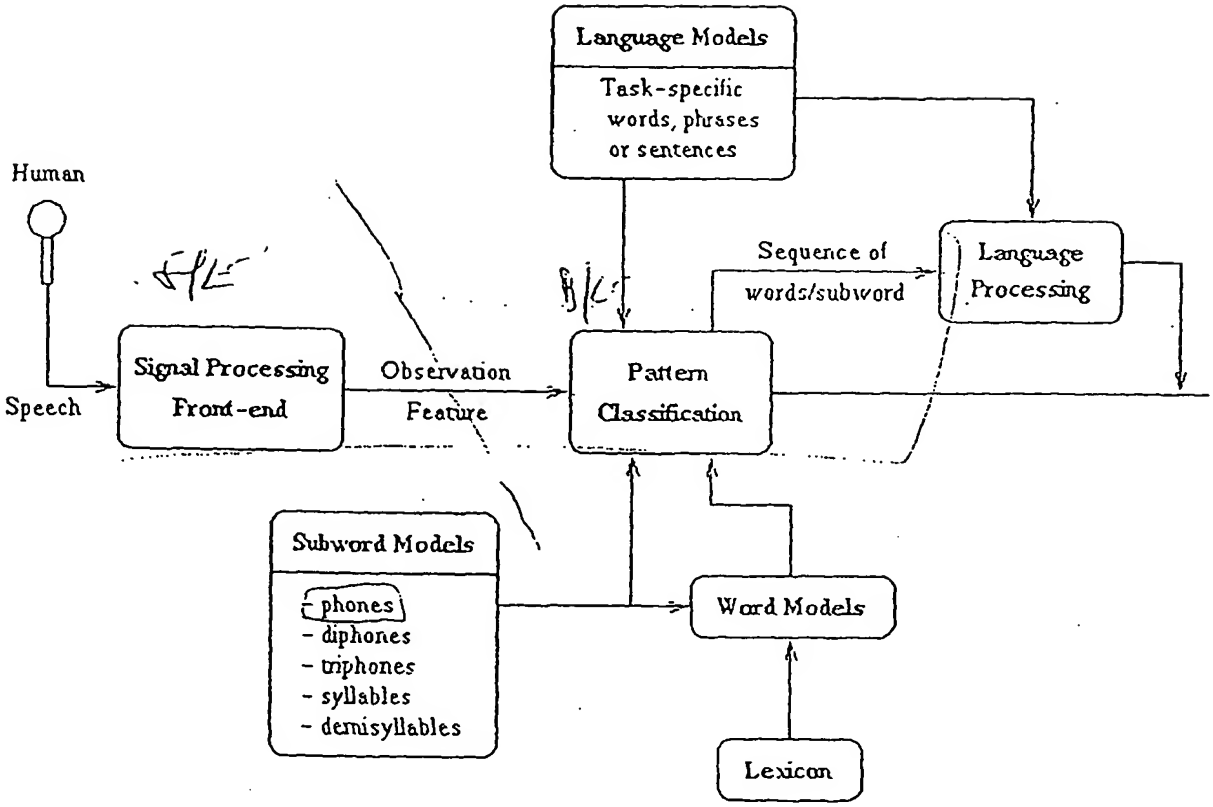

Language Models
Task-specific words, phrases or sentences
Human
Sequence of words/subword
Language Processing
Signal Processing Front-end
Observation Feature
Pattern Classification
Speech
Subword Models
- phones
- diphones
- triphones
- syllables
- demisyllables
Word Models
Lexicon

Fig. 1

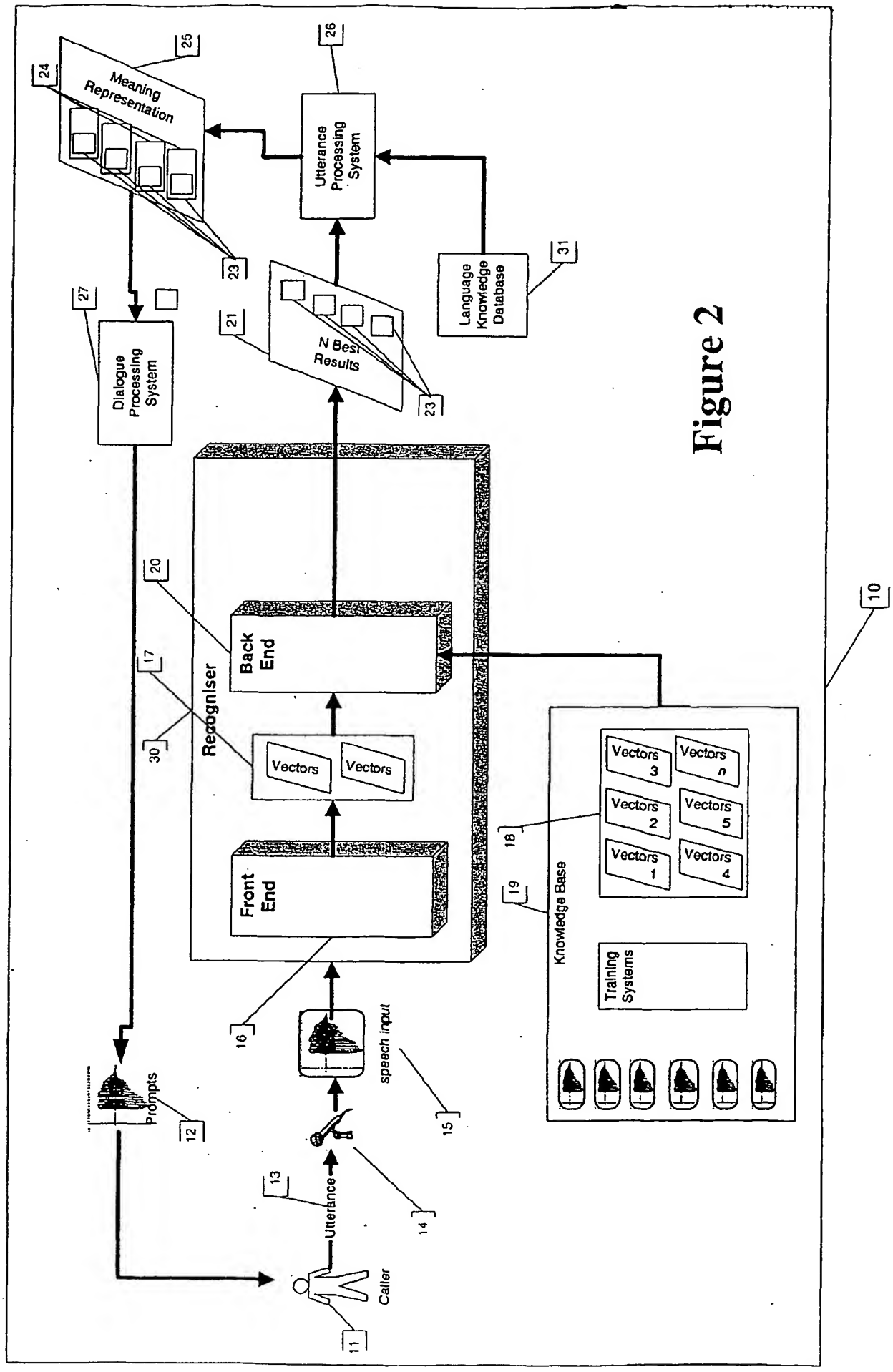
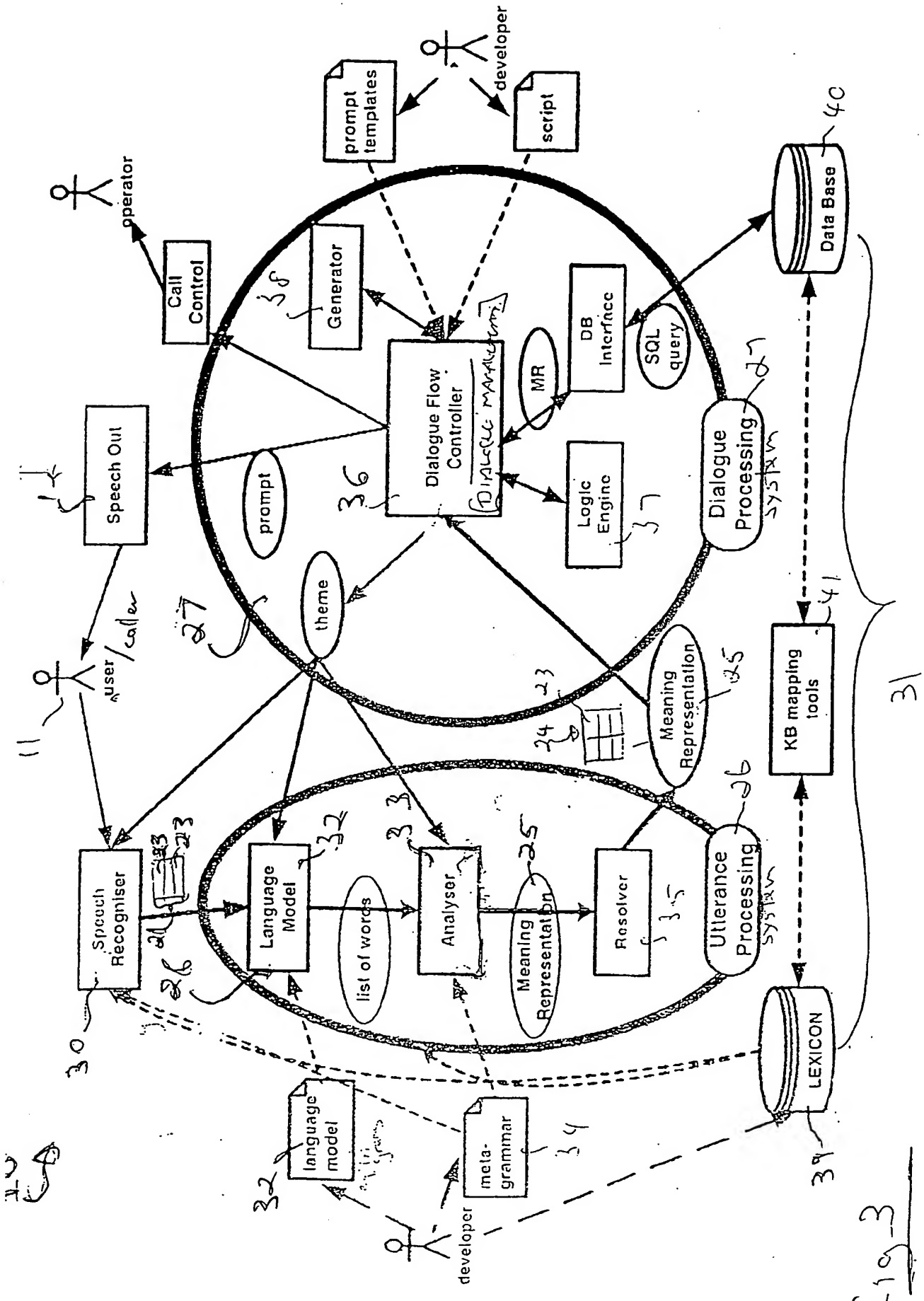
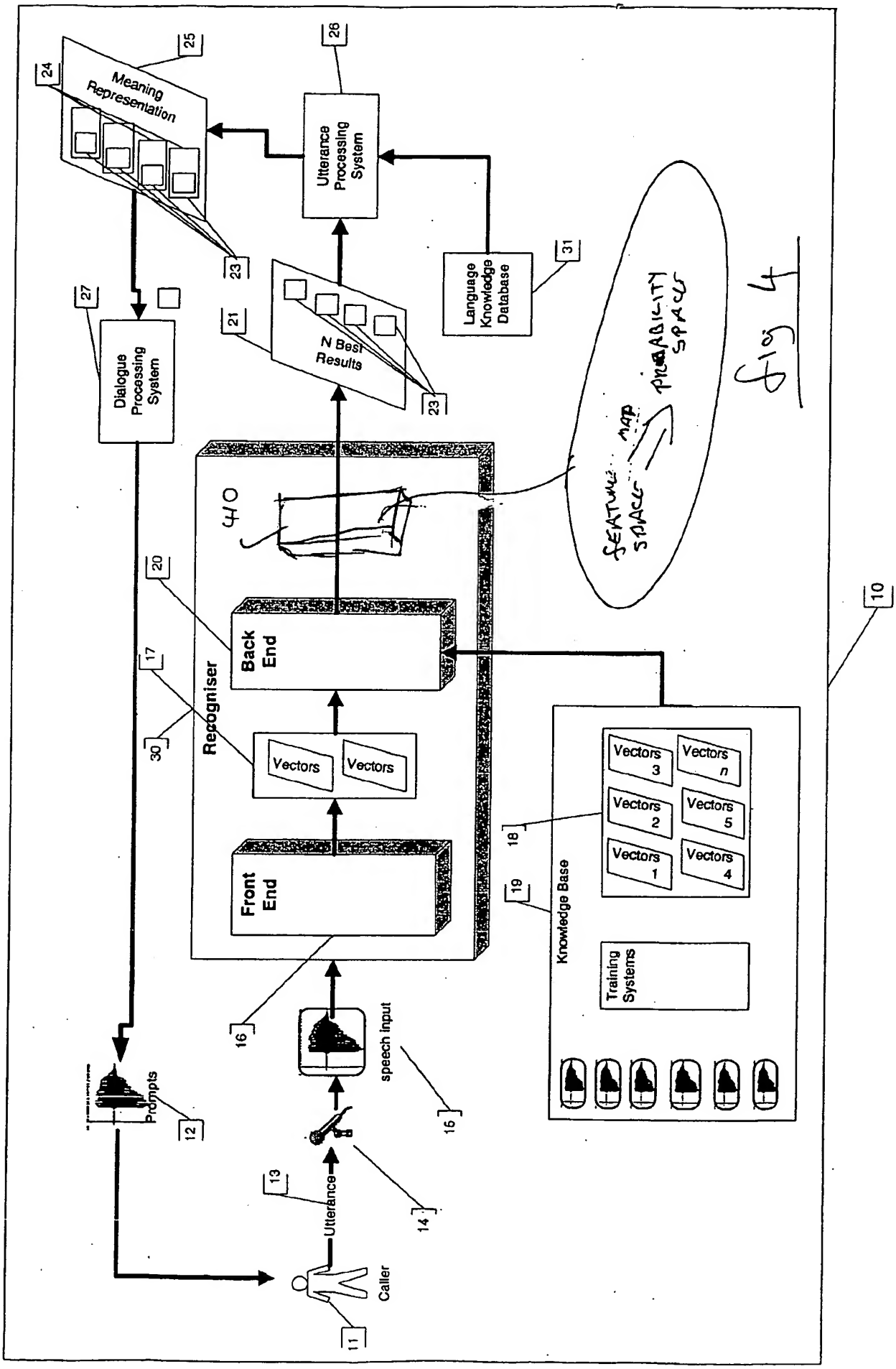

Meaning Representation
Utterance Processing System
Language Knowledge Database
N Best Results
Dialogue Processing System
Recognlser
Back End
Front End
Vectors
Vectors
Knowledge Base
Vectors 3
Vectors n
Vectors 2
Vectors 5
Vectors 1
Vectors 4
Training Systems
Prompts
speech input
Utterance
Caller
10
11
12
13
14
15
16
17
18
19
20
21
23
24
25
26
27
30
31

Figure 2


developer
prompt templates
script
operator
Call Control
Generator
Data Base
DB Interface
SQL query
MR
Dialogue Flow Controller
Logic Engine
Speech Out
prompt
theme
Dialogue Processing system
user / caller
Meaning Representation
KB mapping tools
Speech Recogniser
Language Model
list of words
Analyser
Meaning Representation
Resolver
Utterance Processing system
LEXICON
language model
meta-grammar
developer


Meaning Representation
25
24
26
Utterance Processing System
23
Language Knowledge Database
31
27
Dialogue Processing System
21
N Best Results
23
40
Recogniser
Back End
20
17
30
Vectors
Vectors
Front End
16
Knowledge Base
18
19
Vectors 3
Vectors n
Vectors 2
Vectors 5
Vectors 1
Vectors 4
Training Systems
10
speech input
16
Prompts
12
13
Utterance
14
Caller
11

**INTERNATIONAL SEARCH REPORT**

International application No.
**PCT/AU02/00803**

**A. CLASSIFICATION OF SUBJECT MATTER**

Int. Cl. [7]: G10L 15, G06F 3/16, G06F 17/20, G06F 17/27, G06F 17/28

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
DWPI IPC Marks and Keywords: Speech recognition, Voice recognition, Auto+, Adapt+, Adjust+, Learn+, Teach+, Feedback, Test+, Gaug+, Correlat+, Pars+, Dialog+, Conversat+, Score, Mark, Grade

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| P,X | US 6421641 B (HUANG et al.) 16 July 2002<br>Whole Document | 1-11 |
| X | US 6125345 A (MODI et al.) 26 September 2000<br>Whole Document | 1-11 |
| X | US 6026359 A (YAMAGUCHI et al.) 15 February 2000<br>Whole Document | 1-11 |

[X] Further documents are listed in the continuation of Box C [X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 29 August 2002 | 9 SEP 2002 |
| Name and mailing address of the ISA/AU<br>AUSTRALIAN PATENT OFFICE<br>PO BOX 200, WODEN ACT 2606, AUSTRALIA<br>E-mail address: pct@ipaustralia.gov.au<br>Facsimile No. (02) 6285 3929 | Authorized officer<br>**J.W. THOMSON**<br>Telephone No : (02) 6283-2214 |

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU02/00803

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | US 5465321 A (SMYTH) 7 November 1995<br>Whole Document | 1-11 |

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.
**PCT/AU02/00803**

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

| Patent Document Cited in Search Report | Patent Family Member |
|---|---|
| US 6421641 | NONE |
| US 6125345 | NONE |
| US 6026359 | EP 831461 JP 10149191 |
| US 5465321 | NONE |

END OF ANNEX